

# Cooperative Channel Estimation for Coordinated Transmission with Limited Backhaul

Qianrui Li, *Member, IEEE*, David Gesbert, *Fellow, IEEE*  
and Nicolas Gresset, *Member, IEEE*

## Abstract

Obtaining accurate global channel state information (CSI) at multiple transmitter devices is critical to the performance of many coordinated transmission schemes. Practical CSI local feedback often leads to noisy and partial CSI estimates at each transmitter. With rate-limited bi-directional backhaul, transmitters have the opportunity to exchange a few CSI-related bits towards establishing global CSIT. This work investigates the possible strategies towards this goal. We propose a novel decentralized algorithm that produces MMSE-optimal global channel estimates at each device from combining local feedback and backhaul-exchanged information. The method adapts to arbitrary initial information topologies and feedback noise statistics. The advantage over conventional CSI exchange mechanisms in a network MIMO (CoMP) setting are highlighted.

## Keywords

*Decentralized estimation, Finite-capacity backhaul, Limited feedback, Coordination, Cooperative communication.*

## I. INTRODUCTION

The acquisition of global CSI at multiple devices that engage in the coordinated transmission [1] is crucial to system performance in many cooperative transmission techniques such as network MIMO [2]–[4], coordinated beamforming and scheduling [5]–[7], and interference alignment [8].

---

Part of this work has been presented at the International Conference on Communication (ICCW 2015) Small Cell and 5G Networks (SmallNets) Workshop, London, June 2015.

In order to facilitate cooperation, traditionally, finite-capacity backhaul [9] between transmitters can be exploited among transmitters that need to cooperate. In [9] [10], they have analyzed the multi-cell processing performance with finite-capacity backhaul using information theory tools. The work in [11], [12] also consider the sum rate maximization of the CoMP system with constrained backhaul. A quantization scheme for CSI sharing under finite-capacity backhaul assumption implementing interference alignment is proposed by [13]. However, such designs employ CSI quantizations, with no regard for the *statistical properties* of the local information already existing at the transmitter receiving the information and ignoring the potential benefits of *correlated initial channel estimates* available at the transmitters.

In this work we recast the problem into a more general and systematic decentralized channel estimation problem with side information [14]. In this setup, each transmitter (TX) starts by acquiring an initial CSI estimate from any local feedback mechanism. Interestingly, such mechanisms can be of arbitrary nature, encompassing scenarios such as the current LTE release, where each base station can only acquire CSI related to a subset of the users which are served by that base station. Other scenarios can be also accounted such as broadcast feedback (feedback is overheard by all TX within a certain distance) and hierarchical feedback designs, where in the latter some of the TXs (for e.g. so-called "master base stations") are endowed by design with a greater amount of CSI, compared with surrounding "slave" transmitters [15]. More generally, the initial CSI structure may exhibit an arbitrary level of accuracy as well as spatial correlation (from TX to TX). A general and not previously addressed problem can be then formulated as follows: Given the arbitrary initial CSI structure and given finite information exchange capability between the TXs, what are reasonable strategies for cooperation (among transmitters) for the purpose of generating a high-enough quality CSI at each one of them? An interesting side question is how much information should flow in each direction for every TX pair when backhaul links are subject to a global bidirectional rate constraint.

In this paper, the information exchange through capacity-limited backhaul is modeled via a fixed rate quantization scheme. The final CSI estimate is generated based on minimum mean squared error (MMSE) criterion, and involves a suitable combining of the initial local CSI feedback and the backhaul-exchanged information acquired from other TXs. A difficulty in this problem lies in the fact that the optimization of the information exchange schemes and that of the CSI combining scheme are fundamentally coupled. Nevertheless our contribution reveals that

the two optimization steps can be undertaken jointly.

Clearly the problem of cooperative channel estimation is rooted in the information theoretic framework of network vector quantization [16] and lossy source coding with side information (i.e. Wyner-Ziv coding [17]). When more than two transmitters are involved in the cooperation, it becomes related to the problem of multiple-source compression with side information at the decoder [18]. The information theoretical bound [19], [20] and asymptotically bound-achieving quantizer design for Wyner-Ziv coding [21]–[23] are well analyzed. In this paper we are interested in reasonable complexity, practically implementable optimization algorithms for which Wyner Ziv coding schemes can serve as useful benchmarks.

In this work, we propose a novel optimization framework, referred as *coordination shaping*, which addresses the above problems under a wide range of noise and initial CSI feedback design. Our specific contributions include:

- A joint optimized quantization and information combining scheme allowing to produce MMSE-optimal global CSI at all nodes of the network. The quantizer minimizes a weighted distortion measure where the weight (quantization shaping) matrix is optimized as function of distribution of CSI quality across the cooperating transmitters. The final CSI estimate at each TX linearly combines the initial and exchanged CSI. A key finding is that the quantization shaping matrices and linear combining weights can be optimized jointly by a convex program.
- For the case of two transmitters, our proposed algorithm can even outperforms the Wyner-Ziv transform coding algorithm [23] in the low rate region, while the performance asymptotically achieves the Wyner-Ziv bound in the high resolution regime.
- The proposed algorithm works for multi-transmitter cooperation scenarios as well, hence offering a generalized low-complexity implementation of Wyner-Ziv coding based schemes.
- The proposed framework is exploited to find the optimal coordination bit allocation in the case of global bidirectional rate constraint.

## II. SYSTEM MODEL AND PROBLEM DESCRIPTION

We consider a communication system with  $K$  transmitters and  $L$  receivers (RX). The cooperative TXs could be base stations attempting to serve receiving terminals in a cooperative fashion. There exists a possible cooperative transmission strategies, generally requiring the availability

of some global CSI at each TX [2], [5], [7], [8]. Although the actual choice of the transmission scheme (joint MIMO precoding, interference alignment, coordinated scheduling, coordinated resource allocation, etc.) may affect the CSI reconstruction problem at the TX side, such a question is left for further work while this paper focuses instead on the general problem of producing the best possible global CSI at each and every TX in a non discriminatory manner. The impact of our channel estimation framework on the overall system performance is however evaluated in Section VII for a particular example of network-MIMO enabled system.

Consider each TX  $i, \forall i = 1, \dots, K$  is equipped with  $M$  transmit antennas while each RX  $j, \forall j = 1, \dots, L$  is equipped with  $N$  receive antennas. The propagation channel between TX  $i$  and RX  $j$  is denoted as  $\mathbf{H}_{ji} \in \mathbb{C}^{N \times M}$ . The full network-wide MIMO channel is  $\mathbf{H} \in \mathbb{C}^{NL \times MK}$  with:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \dots & \mathbf{H}_{1K} \\ \vdots & \dots & \vdots \\ \mathbf{H}_{L1} & \dots & \mathbf{H}_{LK} \end{bmatrix}.$$

We consider frequency-flat Rayleigh fading channels, with  $\mathbf{h} = \text{vec}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_h)$ , where  $\mathbf{Q}_h$  is an arbitrary multi-user channel covariance matrix.

#### A. Distributed CSI model

For the CSI model, we assume that each TX acquires an initial estimate of the global channel state from an arbitrary pilot-based, digital or analog feedback mechanism. Similar to the CSI model used in [24], [25], the CSI made initially available at TX  $i$  is a noisy one. More generally, here the CSI imperfection is TX-dependent, giving rise to a distributed CSI model as initially introduced in [26]. Let

$$\hat{\mathbf{H}}^{(i)} = \mathbf{H} + \mathbf{E}^{(i)}, \quad (1)$$

where  $\hat{\mathbf{H}}^{(i)} \in \mathbb{C}^{NL \times MK}$  is a CSI estimate for  $\mathbf{H}$  at TX  $i$ .  $\mathbf{E}^{(i)} \in \mathbb{C}^{NL \times MK}$  is the estimation error seen at TX  $i$ . Hence, the *estimates* at various TXs can be correlated through  $\mathbf{H}$ . The channel independent  $\mathbf{E}^{(i)}$  satisfies  $\mathbf{e}^{(i)} = \text{vec}(\mathbf{E}^{(i)}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_i)$ . The errors terms seen at different TXs are assumed independent, i.e,  $\mathbb{E}\{\mathbf{e}^{(i)} \cdot \mathbf{e}^{(j)H}\} = \mathbf{0}, \forall i \neq j$ . Throughout this work, the channel statistics  $\mathbf{Q}_h$  and all error statistics  $\mathbf{Q}_i, \forall i = 1, \dots, K$  are assumed to be known at every TX by virtue of slow statistical variations.

Note that the value of this CSI model lies in the fact it is quite general, including diverse scenarios ranging from local to global information structures. For example, in the conventional LTE downlink channel estimation scenario, the channel estimation is performed in FDD mode by each Base Station (TX) sending pilots to the users (RXs). Each user will feedback its downlink CSI to its associated Base Station only. This gives rise to a strongly local initial CSI at each TX. In the scenario of *broadcast feedback*, a terminal feeds back its downlink CSI to all overhearing TXs, thus providing global CSI estimates at all TXs where it matters most. Still, in this case too, the quality of the initial CSI estimates remains TX-dependent due to small scale and large scale effects on the uplink. More general various degrees of *locality* can be assigned to the initial CSI structure by selecting a particular  $\mathbf{Q}_i$  with selected matrix components having larger variance than at other TXs. A channel component with corresponding noise variance equal to zero in  $\mathbf{Q}_i$  will indicate perfect knowledge of that coefficient is available at TX  $i$ . Finally, selecting  $\mathbf{Q}_i, \forall i$  to be zero matrices will indicate perfect centralized CSI, which renders further backhaul-based CSI exchange superfluous.

### B. Limited rate coordination model

Consider that the transmitter devices are equipped with rate-limited bi-directional communication links over which they can exchange a finite amount of CSI related information. Note that we only allow a *single shot* of coordination which consumes  $R_{ki}$  bits of communication from TX  $k$  to TX  $i$  for all cooperating  $(k, i)$  pairs *simultaneously*. The problem is now to optimally exploit this coordination capability so as to acquire the best possible global channel estimate at each TX.

In Fig. 1, the cooperation information exchange between two transmitters TX  $i$  and TX  $k$  is illustrated. TX  $k$  sends to TX  $i$  a suitably quantized version of initial CSI estimate  $\hat{\mathbf{h}}^{(k)}$ , denoted by  $\mathbf{z}_{ki}$ . Similar operation is performed by TX  $i$  sending  $\mathbf{z}_{ik}$  to TX  $k$ . The quantization operation associated to the link from TX  $k$  to TX  $i$  is defined as  $\mathcal{Q}_{ki} : \mathbb{C}^n \mapsto \mathcal{C}_{ki}, \mathbf{z}_{ki} \in \mathcal{C}_{ki}, |\mathcal{C}_{ki}| = 2^{R_{ki}}$  where  $\mathcal{C}_{ki}$  is the codebook for the quantizer  $\mathcal{Q}_{ki}$ ,  $n = NMKL$  is the length of quantization vector.

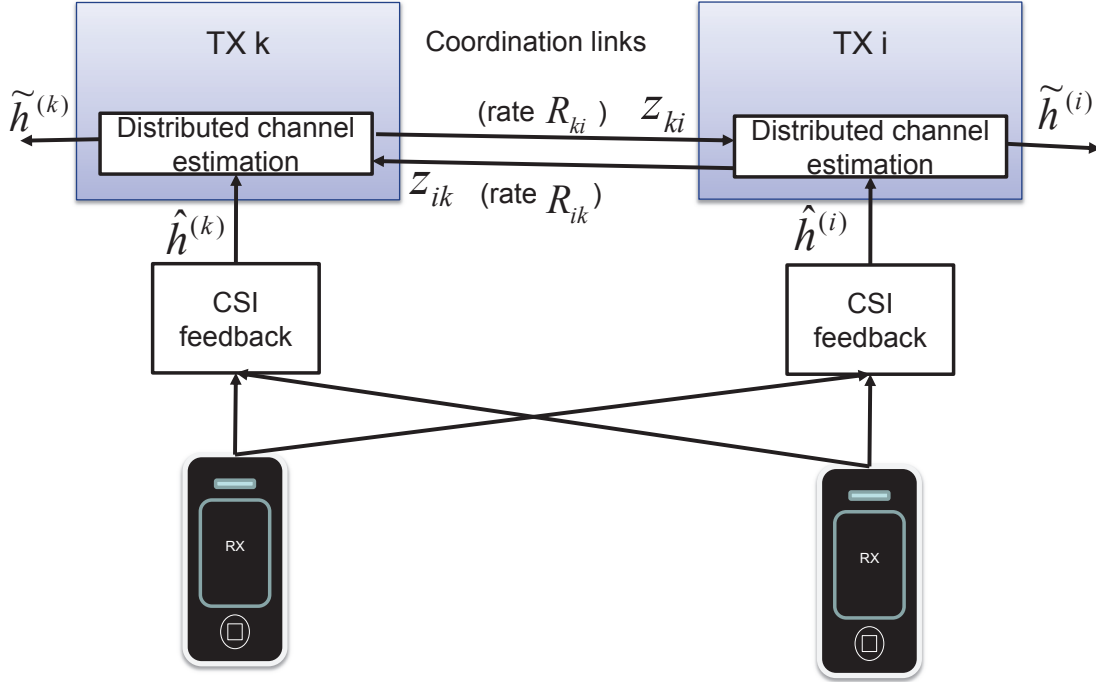


Fig. 1. Decentralized cooperative channel estimation across two TXs serving two terminals.

### C. Channel estimation with limited coordination

At TX  $i$ , a reconstruction function  $g_i(\cdot)$  combines the initial CSI  $\hat{\mathbf{h}}^{(i)}$  and the exchanged CSI  $\mathbf{z}_{ki}$  to form a final estimate  $\tilde{\mathbf{h}}^{(i)}$ .

The MMSE estimation problem at TX  $i$  can be formulated as follows:

$$D^{(i)} = \min \frac{1}{n} \mathbb{E} \{ \|\mathbf{h} - \tilde{\mathbf{h}}^{(i)}\|^2 \} \quad (2)$$

$$= \min_{g_i, \mathcal{Q}_{ki}} \frac{1}{n} \mathbb{E} \{ \|\mathbf{h} - g_i(\hat{\mathbf{h}}^{(i)}, \mathcal{Q}_{ki}(\hat{\mathbf{h}}^{(k)}))\|^2 \}, \quad (3)$$

where  $\tilde{\mathbf{h}}^{(i)} = g_i(\hat{\mathbf{h}}^{(i)}, \mathcal{Q}_{ki}(\hat{\mathbf{h}}^{(k)}))$ .

Note that it is in general a difficult functional optimization and the two functions  $g_i(\cdot)$  and  $\mathcal{Q}_{ki}$  are intertwined.

The goal of this work is to find (i) a suitable reconstruction function  $g_i(\cdot)$  and (ii) the optimal quantizer  $\mathcal{Q}_{ki}$  such that at TX  $i$ ,  $D^{(i)}$  is minimized.

### III. OPTIMAL VECTOR QUANTIZATION MODEL

We now first introduce a useful model for the optimal vector quantization (VQ) which will be exploited in the latter analysis.

Generally, optimal VQ can be derived via a Lloyd-Max algorithm as depicted in Fig. 2. The quantization result  $\mathbf{z}_{ki}$  and quantization error  $\mathbf{e}_{\mathcal{Q}_{ki}}$  are uncorrelated but the quantization input  $\hat{\mathbf{h}}^{(k)}$  is both dependent on  $\mathbf{e}_{\mathcal{Q}_{ki}}$  and  $\mathbf{z}_{ki}$ . The covariance matrices for  $\hat{\mathbf{h}}^{(k)}$ ,  $\mathbf{z}_{ki}$  and  $\mathbf{e}_{\mathcal{Q}_{ki}}$  satisfy [27]

$$\mathbf{Q}_{\hat{\mathbf{h}}^{(k)}} = \mathbf{Q}_{\mathbf{z}_{ki}} + \mathbf{Q}_{\mathbf{e}_{\mathcal{Q}_{ki}}}. \quad (4)$$

Since the input of the quantizer  $\hat{\mathbf{h}}^{(k)}$  is Gaussian, we obtain an upper bound of the impact of quantization by assuming that the quantization error

$$\mathbf{e}_{\mathcal{Q}_{ki}} = \hat{\mathbf{h}}^{(k)} - \mathbf{z}_{ki} \quad (5)$$

is also Gaussian distributed as  $\mathbf{e}_{\mathcal{Q}_{ki}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{\mathcal{Q}_{ki}})$  [28]. Similar to [29] and based on the Gaussian assumption for  $\mathbf{e}_{\mathcal{Q}_{ki}}$ , we can approximate the VQ procedure by a gain-plus-additive-noise model (similar to the scalar quantizer case in [30]) as illustrated in Fig.3.

**Proposition 1.** *Assume the quantization error  $\mathbf{e}_{\mathcal{Q}_{ki}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{\mathcal{Q}_{ki}})$  and is independent from the quantization result  $\mathbf{z}_{ki}$ , the optimal vector quantization for  $\hat{\mathbf{h}}^{(k)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{\mathbf{k}})$  is given by a gain-plus-additive-noise model:*

$$\mathbf{z}_{ki} = (\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{\mathbf{k}} - \mathbf{Q}_{\mathcal{Q}_{ki}})(\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{\mathbf{k}})^{-1}\hat{\mathbf{h}}^{(k)} + \mathbf{q}_{ki}, \quad (6)$$

where  $\mathbf{q}_{ki}$  and  $\hat{\mathbf{h}}^{(k)}$  are uncorrelated random vectors,  $\mathbf{q}_{ki} \sim \mathcal{CN}(\mathbf{0}, (\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{\mathbf{k}} - \mathbf{Q}_{\mathcal{Q}_{ki}})(\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{\mathbf{k}})^{-1}\mathbf{Q}_{\mathcal{Q}_{ki}})$ .

*Proof:* Since  $\mathbf{e}_{\mathcal{Q}_{ki}}$  is assumed to be independent from  $\mathbf{z}_{ki}$ . Knowing that  $\hat{\mathbf{h}}^{(k)} = \mathbf{e}_{\mathcal{Q}_{ki}} + \mathbf{z}_{ki}$ ,  $\mathbf{e}_{\mathcal{Q}_{ki}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{\mathcal{Q}_{ki}})$ ,  $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{\mathbf{h}})$ , according to the Bayesian estimator [31],

$$\begin{aligned} \mathbb{E}\{\mathbf{z}_{ki}|\hat{\mathbf{h}}^{(k)}\} &= (\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{\mathbf{k}} - \mathbf{Q}_{\mathcal{Q}_{ki}})(\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{\mathbf{k}})^{-1}\hat{\mathbf{h}}^{(k)} \\ \text{Cov}\{\mathbf{z}_{ki}|\hat{\mathbf{h}}^{(k)}\} &= (\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{\mathbf{k}} - \mathbf{Q}_{\mathcal{Q}_{ki}})(\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{\mathbf{k}})^{-1}\mathbf{Q}_{\mathcal{Q}_{ki}}, \end{aligned}$$

which concludes the proof. ■

This gain-plus-additive noise model with uncorrelated  $\mathbf{z}_{ki}$  and  $\mathbf{q}_{ki}$  is helpful in the following derivation. In the reminder of this work, both the design of reconstruction functions and optimal quantizers will be based on (6).

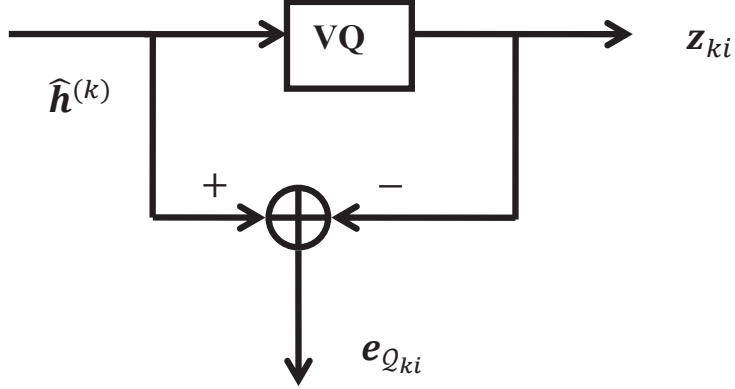


Fig. 2. Quantizer model for optimal vector quantization.

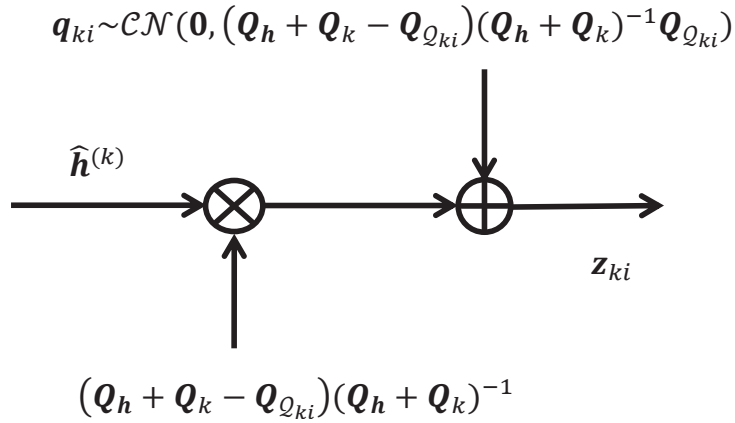


Fig. 3. Gain-plus-additive-noise model for the optimal vector quantization procedure.

#### IV. RECONSTRUCTION FUNCTION DESIGN

This section addresses the aforementioned sub-problem of the optimal reconstruction function design in general settings of multi-TXs cooperation ( $K \geq 2$ ). For ease of illustration, we focus on TX  $i$ , who is cooperating with TX  $k$ ,  $\forall k \in \mathcal{A}_i$ , where the set  $\mathcal{A}_i$  contains the indices of TXs that are cooperating with TX  $i$ .

We consider hereby the reconstruction function as a weighted linear combination of estimates at TX  $i$ , which is suboptimal in a general setting (yet optimal in the particular case of  $K = 2$ ) but leads to a desired closed form optimization.



Hence, the final estimate at TX  $i$  is modeled as:

$$\tilde{\mathbf{h}}^{(i)} = \sum_{k \in \mathcal{A}_i} \mathbf{W}_{ki} \mathbf{z}_{ki} + \mathbf{W}_{ii} \hat{\mathbf{h}}^{(i)}, \quad (7)$$

where  $\mathbf{W}_{ki}, \mathbf{W}_{ii}$  are weighting matrices. The optimal weight combining matrices are revealed now in the following.

**Proposition 2.** *Consider a multi-transmitters cooperation described in (7), assume the CSI estimate at each TX is distributed according to section II-A and the limited rate coordination is modeled according to section II-B, let the quantization error covariance matrices be denoted as  $\mathbf{Q}_{\mathcal{Q}_{ki}}, \forall k \in \mathcal{A}_i$ . The optimum per dimensional MSE for the final estimate at TX  $i$  is:*

$$D^{(i)opt} = \frac{1}{n} \text{Tr} \left( \mathbf{Q}_{\mathbf{h}}^{-1} + \mathbf{Q}_i^{-1} + \mathbf{\Lambda}_i \right)^{-1}, \quad (8)$$

where  $\mathbf{\Lambda}_i$  is defined as:

$$\mathbf{\Lambda}_i = \sum_{k \in \mathcal{A}_i} \left( (\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_k) (\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_k - \mathbf{Q}_{\mathcal{Q}_k})^{-1} (\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_k) - \mathbf{Q}_{\mathbf{h}} \right)^{-1}.$$

The optimal weight combining matrices  $\{\mathbf{W}_{ki}^{opt}, \mathbf{W}_{ii}^{opt}, \forall k \in \mathcal{A}_i\}$  are obtained as:

$$\begin{bmatrix} \mathbf{W}_{ii}^{opt} & \mathbf{W}_{l_1 i}^{opt} & \dots & \mathbf{W}_{l_{C_i} i}^{opt} \end{bmatrix} = \mathbf{Q}_{\mathbf{h}} \mathbf{\Upsilon}_i \mathbf{\Omega}_i^{-1}, \quad (9)$$

where  $\mathbf{\Upsilon}_i, \mathbf{\Omega}_i$  are given below, the set  $\mathcal{A}_i$  has cardinality  $|\mathcal{A}_i| = C_i$  and each element in set is denoted by  $\mathcal{A}_i = \{l_1, \dots, l_{C_i}\}$ .

$$\begin{aligned} \mathbf{P}_{l_j i} &= \mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{l_j} - \mathbf{Q}_{\mathcal{Q}_{l_j i}} \\ \mathbf{A}_{l_j i} &= \mathbf{P}_{l_j i} (\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_{l_j})^{-1}, \quad \forall j = 1 \dots C_i \\ \mathbf{\Upsilon}_i &= \begin{bmatrix} \mathbf{I} & \mathbf{A}_{l_1 i}^H & \dots & \mathbf{A}_{l_{C_i} i}^H \end{bmatrix} \\ \mathbf{\Omega}_i &= \begin{bmatrix} \mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_i & \mathbf{Q}_{\mathbf{h}} \mathbf{A}_{l_1 i}^H & \dots & \mathbf{Q}_{\mathbf{h}} \mathbf{A}_{l_{C_i} i}^H \\ \mathbf{A}_{l_1 i} \mathbf{Q}_{\mathbf{h}} & \mathbf{P}_{l_1 i} & \dots & \mathbf{A}_{l_1 i} \mathbf{Q}_{\mathbf{h}} \mathbf{A}_{l_{C_i} i}^H \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{l_{C_i} i} \mathbf{Q}_{\mathbf{h}} & \mathbf{A}_{l_{C_i} i} \mathbf{Q}_{\mathbf{h}} \mathbf{A}_{l_1 i}^H & \dots & \mathbf{P}_{l_{C_i} i} \end{bmatrix}. \end{aligned}$$

*Proof:* See Appendix A. ■

*Remark 1.* The optimal weight combining matrices  $\{\mathbf{W}_{ki}^{opt}, \mathbf{W}_{ii}^{opt}, \forall k \in \mathcal{A}_i\}$  and  $D^{(i)opt}$  are merely functions of statistics  $\mathbf{Q}_{\mathbf{h}}, \mathbf{Q}_i, \mathbf{Q}_k, \mathbf{Q}_{\mathcal{Q}_{ki}}, \forall k \in \mathcal{A}_i$ . □

*Remark 2.* Consider a motivation example of two TXs cooperation, at TX 1, the final estimate is:

$$\tilde{\mathbf{h}}^{(1)} = \mathbf{W}_{21}\mathbf{z}_{21} + \mathbf{W}_{11}\hat{\mathbf{h}}^{(1)},$$

where

$$\begin{aligned} \mathbf{P}_{21} &= \mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_2 - \mathbf{Q}_{\mathbf{Q}_{21}} \\ \mathbf{A}_{21} &= \mathbf{P}_{21}(\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_2)^{-1} \\ [\mathbf{W}_{11}, \mathbf{W}_{21}] &= \mathbf{Q}_{\mathbf{h}} \begin{bmatrix} \mathbf{I} & \mathbf{A}_{21}^H \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_1 & \mathbf{Q}_{\mathbf{h}}\mathbf{A}_{21}^H \\ \mathbf{A}_{21}\mathbf{Q}_{\mathbf{h}} & \mathbf{P}_{21} \end{bmatrix}^{-1}. \end{aligned}$$

The optimal per dimensional MSE is:

$$\begin{aligned} D^{(1)opt} &= \frac{1}{n} \text{Tr}(\mathbf{Q}_{\mathbf{h}}^{-1} + \mathbf{Q}_1^{-1} + \mathbf{\Lambda}_1)^{-1} \\ \mathbf{\Lambda}_1 &= ((\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_2)\mathbf{P}_{21}^{-1}(\mathbf{Q}_{\mathbf{h}} + \mathbf{Q}_2) - \mathbf{Q}_{\mathbf{h}})^{-1}. \end{aligned}$$

□

*Remark 3.* The optimal per dimensional MSE and the error covariance matrix for the final estimate is related to 3 covariance terms:  $\mathbf{Q}_{\mathbf{h}}$  indicates the intrinsic (true) channel statistics,  $\mathbf{Q}_i$  refers to the initial estimation error covariance and  $\mathbf{\Lambda}_i$  is related to the initial estimation error and the quantization error covariance at all TXs that cooperate with TX  $i$ . The covariance of the final estimate is formulated as the inverse of the sum of the 3 terms' individual inverse. □

## V. QUANTIZER DESIGN

For the optimal quantizer design, it should be noticed that a conventional optimal VQ (optimal VQ with MSE distortion) implemented by Lloyd-Max algorithm is far from being optimal because rather than minimizing the per dimensional MSE  $D^{(i)}$  for the final estimate, it only guarantees that the quantization distortion will be minimized.

Therefore, the quantizer should be properly shaped such that the quantization procedure ensures not only the minimization of quantization distortion, but also guarantees that the quantization result, after weighted combination with other estimates, will have the minimal per dimensional MSE  $D^{(i)}$ . A useful interpretation of this approach is as follows. As  $\mathbf{Q}_i, \forall i = 1 \dots K$  reflect the spatial distribution of accuracy of the initial CSI, the quantizer  $\mathcal{Q}_{ki}$  should allocate the quantization resource where more bits are needed, i.e. in channel elements or directions that

are well known by TX  $k$  and least known by TX  $i$ . To this end, we choose the weighted square error distortion

$$d_{\mathcal{Q}_{ki}}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^H \mathbf{B}_{ki} (\mathbf{x} - \mathbf{y}) \quad (10)$$

as the distortion measure of the quantizer  $\mathcal{Q}_{ki}$  with the positive definite shaping matrix  $\mathbf{B}_{ki}$  to be optimized.

For an important intermediate step, we can calculate  $\mathbf{Q}_{\mathcal{Q}_{ki}}$  in the asymptotic case as a function of  $\mathbf{B}_{ki}$  when the given coordination link rate  $R_{ki}$  is sufficiently large, i.e. in the high resolution regime.

**Proposition 3.** *Consider quantization on a  $n$  dimensional complex random vector source  $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Gamma})$ , in the high resolution regime where the number of quantization level  $S$  is large, the optimal VQ using weighted MSE distortion with shaping matrix  $\mathbf{B}$  will have a quantization error covariance matrix  $\mathbf{Q}_{\mathbf{x}}$  as:*

$$\mathbf{Q}_{\mathbf{x}} = \mathbf{Q}_0^{(S)}(\mathbf{\Gamma}) \det(\mathbf{B})^{\frac{1}{n}} \mathbf{B}^{-1},$$

where

$$\mathbf{Q}_0^{(S)}(\mathbf{\Gamma}) = S^{-\frac{1}{n}} M_{2n} 2\pi \left( \frac{n+1}{n} \right)^{n+1} \det(\mathbf{\Gamma})^{\frac{1}{n}} \mathbf{I}_n$$

is the quantization error covariance matrix for  $\mathbf{x}$  in the high resolution regime when conventional optimal VQ is applied [32].  $M_{2n}$  is a constant related to  $2n$ .

*Proof:* See Appendix B. ■

*Remark 4.* The aforementioned quantization error covariance matrix expression encompasses the quantization error covariance matrix for conventional optimal VQ by taking  $\mathbf{B} = \mathbf{I}_n$ . It can be easily verified that by imposing a constraint that  $\det(\mathbf{B}) = 1$ , for all values of matrix  $\mathbf{B}$ , the corresponding quantizers will have the same quantization distortion. □

*Remark 5.* For the constant  $M_{2n}$ , a look-up table for  $2n = 1, \dots, 10$  in [33] can be used. when  $n$  is larger, we can approximate  $M_{2n} = \frac{1}{2\pi e}$ . □

We now exploit Proposition 3 in order to derive  $\mathbf{Q}_{\mathcal{Q}_{ki}}$ :

$$\mathbf{Q}_{\mathcal{Q}_{ki}} = \mathbf{Q}_0^{(S)}(\mathbf{\Gamma}) \det(\mathbf{B}_{ki})^{\frac{1}{n}} \mathbf{B}_{ki}^{-1}, \quad (11)$$

where

$$\begin{aligned} S &= 2^{R_{ki}} \\ \Gamma &= \mathbf{Q}_h + \mathbf{Q}_k. \end{aligned}$$

#### A. Shaping matrix optimization

Based on the reconstruction function in Section IV and using equations (8), (11), we can now proceed with the task of jointly optimizing the reconstruction function and the quantizer by solely optimizing the value of  $\mathbf{B}_{ki}$ .

$$\begin{aligned} \min_{\mathbf{B}_{ki}, k \in \mathcal{A}_i} \quad & D^{(i)opt} \\ \text{s.t.} \quad & \det(\mathbf{B}_{ki}) = 1, \mathbf{B}_{ki} \succeq 0 \\ & D^{(i)opt} \text{ defined in (8)}. \end{aligned} \tag{12}$$

As is mentioned in Remark 4, the constraints on  $\mathbf{B}_{ki}$  matrices ensure that all feasible quantizers have the same quantization distortion, the optimization will find  $\mathbf{B}_{ki}, \forall k \in \mathcal{A}_i$  that minimize the per dimensional MSE for the final estimate.

*Proposition 4.* The objective function in problem (12) is convex. In high resolution regime, problem (12) can be approximated by the following convex optimization problem:

$$\begin{aligned} \min_{\mathbf{B}_{ki}, k \in \mathcal{A}_i} \quad & \frac{1}{n} \text{Tr} \left( \sum_{k \in \mathcal{A}_i} (\mathbf{Q}_k^{-1} - \mathbf{Q}_k^{-1} \mathbf{Q}_{\mathcal{Q}_{ki}} \mathbf{Q}_k^{-1}) + \mathbf{Q}_i^{-1} + \mathbf{Q}_h^{-1} \right)^{-1} \\ \text{s.t.} \quad & \det(\mathbf{B}_{ki}) \geq 1, \mathbf{B}_{ki} \succeq 0 \\ & \mathbf{Q}_{\mathcal{Q}_{ki}} \text{ defined in (11)}. \end{aligned} \tag{13}$$

*Proof:* See Appendix C. ■

The reason for solving optimization problem (13) rather than solving directly the original optimization problem (12) is that the former can be easily transformed into a semi-definite quadratic linear programming. It can be solved efficiently by optimization toolbox such as CVX. It should be noted that once the optimal weight matrix  $\mathbf{B}_{ki}^*$  is obtained, the codebook for optimal quantizer  $\mathcal{Q}_{ki}^*$  can be calculated based on Lloyd algorithm and a training set. The optimal weight matrices for estimation combine can be calculated according to (9). Noticing that this optimization is semi-static, the weight matrices for the estimates combination and the shaping matrices for the quantizers will be updated only when the channel statistics or the backhaul resources have been changed.

Interestingly, the asymptotic performance of the proposed algorithm can be characterized in relation to known information theoretic bound.

*Proposition 5.* For a two TXs cooperation scenario of TX 1 and TX 2 as described in section II-C, at TX 1 the proposed coordination shaping algorithm can achieve asymptotically in high resolution regime the Wyner-Ziv bound given by

$$D_{\infty}^{(1)opt} = D_{\infty}^{(1)NWZ} = \frac{1}{n} \text{Tr} (\mathbf{Q}_1^{-1} + \mathbf{Q}_2^{-1} + \mathbf{Q}_h^{-1})^{-1}.$$

*Proof:* the asymptotical per dimension MSE for proposed algorithm is:

$$\begin{aligned} D_{\infty}^{(1)opt} &= \lim_{R_{21} \rightarrow \infty} D^{(1)opt} \\ &= \frac{1}{n} \text{Tr} (\mathbf{Q}_1^{-1} + \mathbf{Q}_2^{-1} + \mathbf{Q}_h^{-1})^{-1}. \end{aligned}$$

It is well known that the information theoretic bound of the per dimension MSE for two TXs cooperation can be achieved using a Wyner-Ziv quantizer and the asymptotic distortion is [23]:

$$D_{\infty}^{(1)NWZ} = \frac{1}{n} \text{Tr}(\mathbb{E}_{YZ} \text{Var}[X|Y, Z]),$$

where  $X, Y, Z$  correspond to the source data, side information and noisy source (i.e, perfect CSI  $\mathbf{h}$ , initial CSI  $\hat{\mathbf{h}}^{(1)}$  and initial CSI at it's cooperation TX  $\hat{\mathbf{h}}^{(2)}$  as in our case). Since Gaussianity is assumed for the perfect CSI and initial CSI,  $D_{\infty}^{(1)NWZ}$  can be calculated as:

$$\begin{aligned} D_{\infty}^{(1)NWZ} &= \frac{1}{n} \text{Tr} \left( \mathbf{Q}_h - \begin{bmatrix} \mathbf{Q}_h & \mathbf{Q}_h \end{bmatrix} \begin{bmatrix} \mathbf{Q}_h + \mathbf{Q}_1 & \mathbf{Q}_h \\ \mathbf{Q}_h & \mathbf{Q}_h + \mathbf{Q}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Q}_h \\ \mathbf{Q}_h \end{bmatrix} \right) \\ &= \frac{1}{n} \text{Tr} (\mathbf{T}_1 - \mathbf{T}_1(\mathbf{T}_1 + \mathbf{Q}_2)^{-1}\mathbf{T}_1) \\ &= \frac{1}{n} \text{Tr} (\mathbf{T}_1^{-1} + \mathbf{Q}_2^{-1})^{-1} \\ &= \frac{1}{n} \text{Tr} (\mathbf{Q}_1^{-1} + \mathbf{Q}_2^{-1} + \mathbf{Q}_h^{-1})^{-1} \\ &= D_{\infty}^{(1)opt}, \end{aligned}$$

where  $\mathbf{T}_1 = \mathbf{Q}_h(\mathbf{Q}_h + \mathbf{Q}_1)^{-1}\mathbf{Q}_1$ . ■

Thus, it reveals that in two TXs cooperation case, the proposed coordination shaping algorithm is asymptotically optimal.

## VI. COORDINATION LINK BIT ALLOCATION

An interesting consequence of the above analysis is the optimization of coordination where multiple transmitters can exchange simultaneously CSI-related information to each other under a global constraint on the coordination bits. The global optimization problem over all coordination links now becomes:

$$\begin{aligned}
 & \min_{\substack{\mathbf{B}_{ki}, R_{ki} \\ k \in \mathcal{A}_i, i=1, \dots, K}} \quad \frac{1}{K} \sum_{i=1}^K D^{(i)opt} \\
 & \text{s.t.} \quad \det(\mathbf{B}_{ki}) = 1, \mathbf{B}_{ki} \succeq 0 \\
 & \quad \sum_{i=1}^K \sum_{k \in \mathcal{A}_i} R_{ki} = R_{tot}, R_{ki} \in \mathbb{N}^+ \\
 & \quad D^{(i)opt} \text{ defined in (8).}
 \end{aligned} \tag{14}$$

Due to the integer constraints on  $R_{ki}$ , this problem becomes a non-convex optimization. However, conventional alternating algorithms can be applied to perform the optimization in a two-step iterative approach.

---

**Algorithm 1** Iterative algorithm for problem 14

---

- 1: Initialize  $\mathbf{B}_{ki}, \forall k \in \mathcal{A}_i, \forall i = 1, \dots, K$
  - 2: **while** not converge **do**
  - 3:     Optimize  $R_{ki}$  with  $\mathbf{B}_{ki}$  fixed
  - 4:     Optimize  $\mathbf{B}_{ki}$  with  $R_{ki}$  fixed
  - 5: **end while**
- 

The optimization in the first step is an integer programming problem and the optimization in the second step is a convex optimization problem. Hence, many conventional algorithms can be applied in both steps. It should be noted that the alternating algorithm does not guarantee the global optimum. Based on the initial point, it might converge to a local optimal point as well.

## VII. NUMERICAL PERFORMANCE ANALYSIS

In this section, a network MIMO transmission [2] setup is considered. Unless otherwise indicated, the default simulation settings are  $K = 2$  and  $L = 2$ ,  $M = N = 1$ . The channel  $\mathbf{h} \in \mathbb{C}^{4 \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_4)$  and the rates on coordination link from TX 2 to TX 1 and from TX 1 to TX 2 are denoted  $R_{21}$  and  $R_{12}$ , respectively. Each TX constructs a ZF precoder based

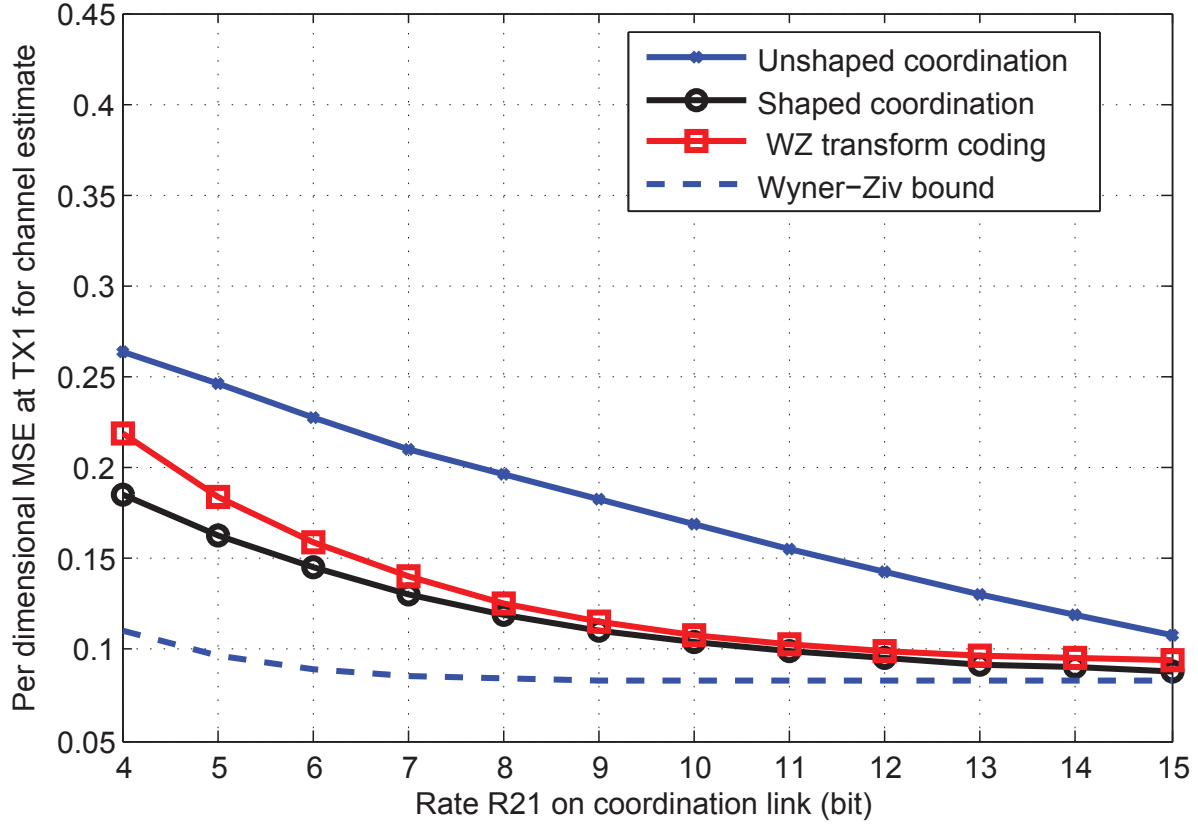


Fig. 4. Mean square error for the final channel estimation at TX 1 as a function of coordination link rate  $R_{21}$ , CSI error covariance matrices are  $\mathbf{Q}_1 = \text{diag}(0.1, 0.1, 0.9, 0.9)$ ,  $\mathbf{Q}_2 = \text{diag}(0.9, 0.9, 0.1, 0.1)$ .

on its final channel estimate. The power control at each TX is 20dB. The per dimensional MSE for decentralized channel estimation is evaluated for different settings using Monte-Carlo simulations over  $10^5$  channel realizations. Since  $n = 2MNKL = 8$ , the parameter  $M_n$  is chosen to be 929/12960 which is related to the  $E8$  lattice [33]. In Fig. 4 and Fig. 6, the CSI information structure is characterized by  $\mathbf{Q}_1 = \text{diag}(0.1, 0.1, 0.9, 0.9)$  and  $\mathbf{Q}_2 = \text{diag}(0.9, 0.9, 0.1, 0.1)$  which corresponds to an example where TX 1 has more accurate CSI about RX 1 and less accurate CSI about RX 2, while TX 2 has more accurate CSI about RX 2 and less accurate CSI about RX 1. The  $\text{diag}(\cdot)$  operator represents a diagonal matrix with diagonal elements in the parenthesis.

Fig.4 shows the per dimensional MSE for the final estimation at TX 1. The Wyner-Ziv bound is the information theoretic bound. The shaped coordination curve applies the proposed algorithm. The unshaped coordination implements the traditional optimal VQ and finds  $\mathbf{W}_{21}, \mathbf{W}_{11}$

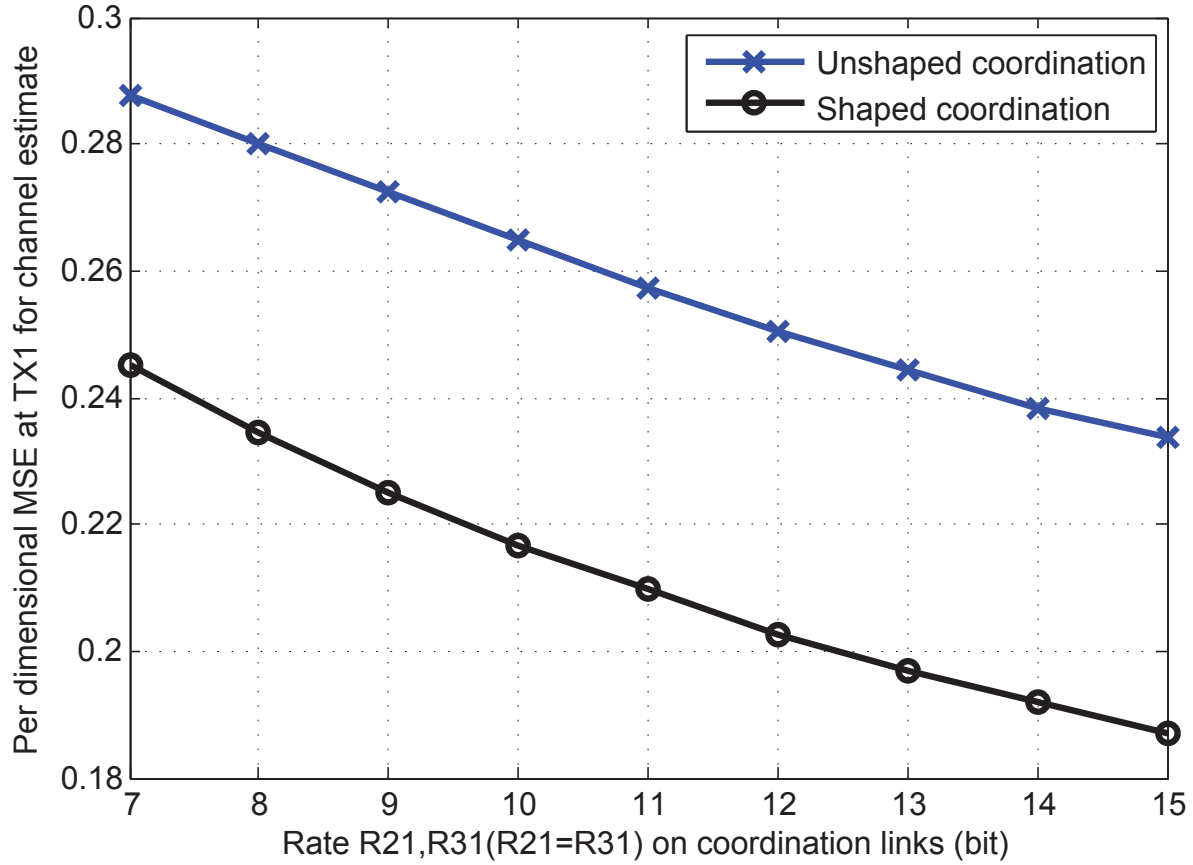


Fig. 5. 3 TX cooperation: Mean square error for the final channel estimation at TX 1 as a function of coordination link rate  $R_{21}(=R_{31})$ .

accordingly using (9). From the figure we can conclude that the shaped coordination algorithm outperforms the unshaped coordination algorithm, which not surprisingly shows the benefit of taking the priori statistic information into account. The WZ transform coding curve refers to the asymptotic optimal Wyner-Ziv transform coding for noisy source in [23]. It reveals that our algorithm outperforms the Wyner-Ziv transform coding algorithm in low coordination rate region and converges asymptotically to the Wyner-Ziv bound  $D_\infty$  as expected.

In Fig.5, the per dimensional MSE for the final estimation at TX 1 is plotted for a 3 TX cooperation network. The rate constraint on coordination link from TX 2 to TX 1 and TX 3 to



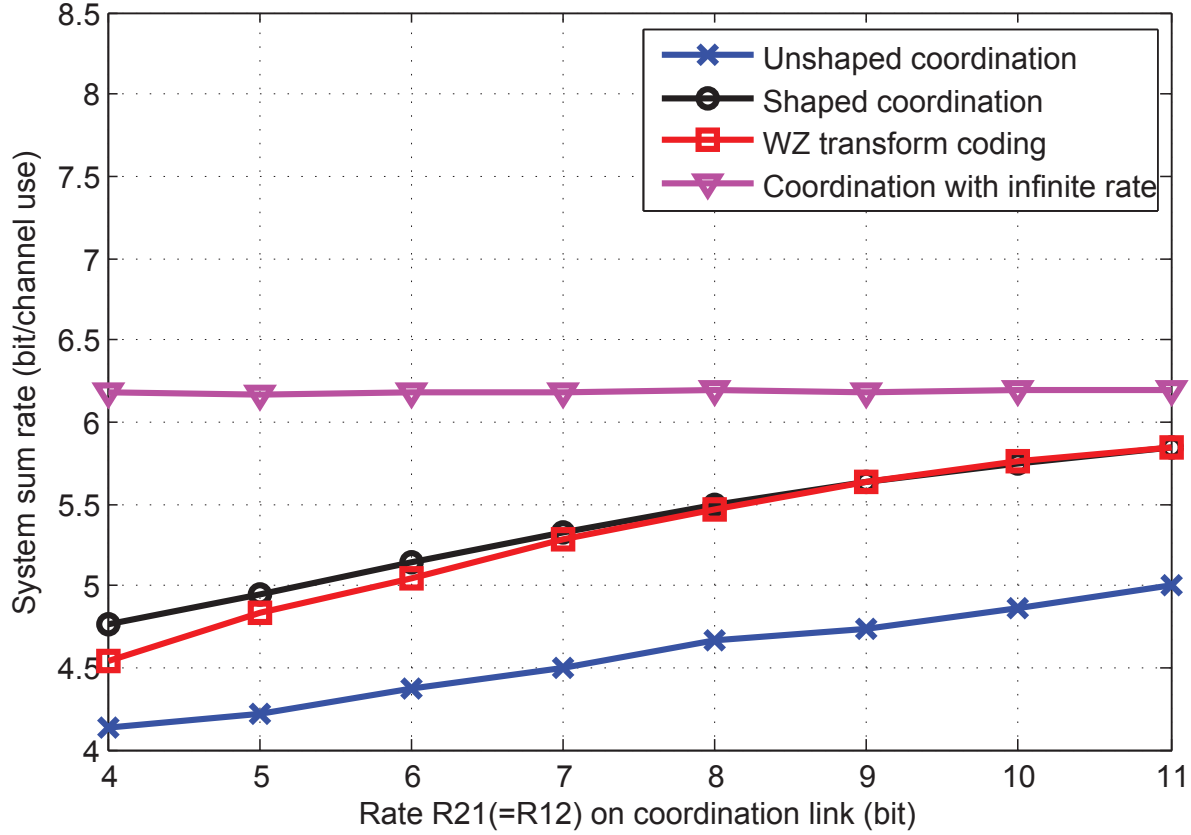


Fig. 6. Sum rate for the 2 TX cooperation system, each TX implements a ZF precoder based on its final channel estimation, SNR is 20dB per TX, CSI error covariance matrices are  $\mathbf{Q}_1 = \text{diag}(0.1, 0.1, 0.9, 0.9)$ ,  $\mathbf{Q}_2 = \text{diag}(0.9, 0.9, 0.1, 0.1)$ .

TX 1 are  $R_{21}$  and  $R_{31}$  respectively. In this simulation, the parameters are denoted below:

$$K = L = 3, M = N = 1, n = 18, M_{2n} = \frac{1}{2\pi e}$$

$$\mathbf{Q}_1 = \text{diag}(0.8147, 0.9058, 0.1270, 0.9134, 0.6324, 0.0975, 0.2785, 0.5469, 0.9575)$$

$$\mathbf{Q}_2 = \text{diag}(0.9649, 0.1576, 0.9706, 0.9572, 0.4854, 0.8003, 0.1419, 0.4218, 0.9157)$$

$$\mathbf{Q}_3 = \text{diag}(0.7922, 0.9595, 0.6557, 0.0357, 0.8491, 0.9340, 0.6787, 0.7577, 0.7431).$$

The simulation clearly shows the performance enhancement of the coordination shaping algorithm over the unshaped coordination algorithm in a multiple TX cooperation scenario.

Fig. 6 exhibits the sum rate for a 2 TX cooperation system. The rate on the coordination links satisfies  $R_{21} = R_{12}$ . We also provide the sum rate for the case when coordination links have infinite bandwidth. The figure shows that the proposed shaped coordination algorithm will

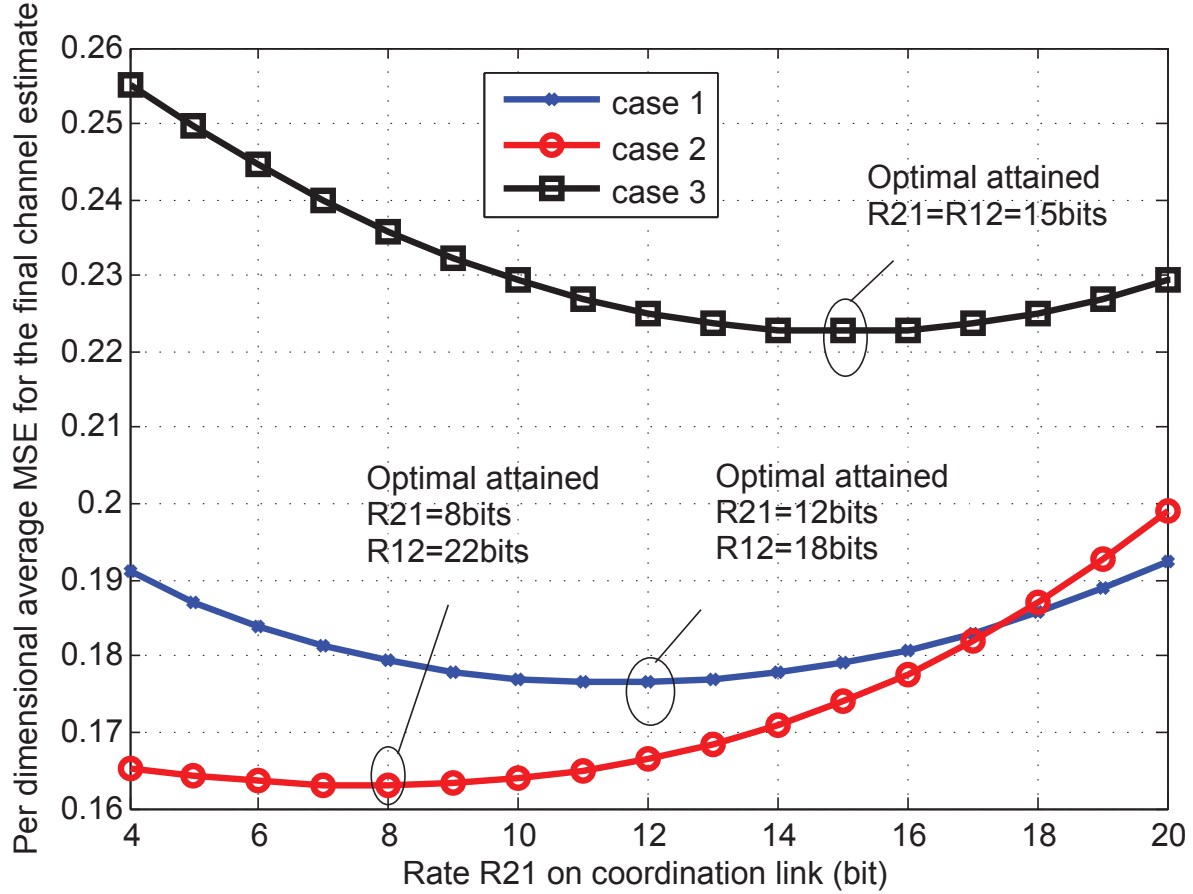


Fig. 7. Per dimensional average mean square error for the final channel estimation at TX 1, TX 2 when sum bit for coordination link  $R_{tot} = R_{21} + R_{12} = 30$ bits. Case 1:  $\mathbf{Q}_1 = \text{diag}(0.1, 0.1, 0.9, 0.9)$ ,  $\mathbf{Q}_2 = \text{diag}(0.5, 0.5, 0.5, 0.5)$ , case 2:  $\mathbf{Q}_1 = \text{diag}(0.4, 0.2, 0.3, 0.1)$ ,  $\mathbf{Q}_2 = \text{diag}(0.7, 0.8, 0.6, 0.9)$ , case 3:  $\mathbf{Q}_1 = \text{diag}(0.5, 0.5, 0.5, 0.5)$ ,  $\mathbf{Q}_2 = \text{diag}(0.5, 0.5, 0.5, 0.5)$ .

improve the system sum rate beyond the unshaped coordination algorithm and WZ transform coding algorithm when a simple ZF precoder is implemented. As the rate on coordination link increases, the sum rate for all algorithms will converge to the infinite coordination rate case.

Fig.7 considers the coordination link bit allocation problem for a 2 TX cooperation system. The total amount of bits for coordination link is  $R_{12} + R_{21} = 30$ bits. The figure reveals that the cooperation information exchange is not necessarily symmetric. In case 3, the two TXs exchange information with equal rate  $R_{12} = R_{21} = 15$ bits because the accuracy of CSI at both end is the same. However, in case 2, the optimal coordination link bit allocation strategy is to let TX 1 share the cooperation information to TX 2 through a  $R_{21} = 8$ bits coordination link and vice versa

through a  $R_{12} = 22$ bits coordination link. It's intuitive because for every channel coefficient, TX 1 has a more accurate initial CSI than TX 2. It reveals that if one TX has a better CSI, it is more encouraged to share his information through a higher rate coordination link.

### VIII. CONCLUSION

We study the decentralized cooperative channel estimation for coordinated transmission with limited backhaul. We derive a low-complexity algorithm which exploits the finite-capacity backhaul near-optimally and is robust to arbitrary feedback noise statistics. We exhibit clear advantages over CSI acquisition and exploitation methods used in conventional CoMP systems.

### APPENDIX A

#### PROOF OF PROPOSITION 2

Adopt the notation in Proposition 2, according to (2), (6) and (7), the per dimensional MSE can be expressed as:

$$\begin{aligned} D^{(i)} &= \frac{1}{n} \mathbb{E}\{\|\mathbf{h} - \tilde{\mathbf{h}}^{(i)}\|^2\} \\ &= \frac{1}{n} \text{Tr} \left( \left( \sum_{k \in \mathcal{A}_i} \mathbf{W}_{ki} \mathbf{A}_{ki} + \mathbf{W}_{ii} - \mathbf{I} \right) \mathbf{Q}_h \left( \sum_{k \in \mathcal{A}_i} \mathbf{W}_{ki} \mathbf{A}_{ki} + \mathbf{W}_{ii} - \mathbf{I} \right)^H \right) \\ &\quad + \frac{1}{n} \text{Tr} \left( \sum_{k \in \mathcal{A}_i} \mathbf{W}_{ki} \left( \mathbf{A}_{ki} \mathbf{Q}_k \mathbf{A}_{ki}^H + \mathbf{A}_{ki} \mathbf{Q}_{\mathcal{Q}_{ki}} \right) \mathbf{W}_{ki}^H + \mathbf{W}_{ii} \mathbf{Q}_i \mathbf{W}_{ii}^H \right). \end{aligned}$$

Take the partial derivatives and set them to zero:

$$\begin{aligned} \frac{\partial D^{(i)}}{\partial \mathbf{W}_{ii}^*} &= 0 \\ \frac{\partial D^{(i)}}{\partial \mathbf{W}_{ki}^*} &= 0, \quad \forall k \in \mathcal{A}_i, \end{aligned}$$

which leads to:

$$\begin{aligned} \mathbf{W}_{ii} (\mathbf{Q}_h + \mathbf{Q}_i) &= \left( \mathbf{I} - \sum_{k \in \mathcal{A}_i} \mathbf{W}_{ki} \mathbf{A}_{ki} \right) \mathbf{Q}_h \\ \mathbf{W}_{ki} \left( \mathbf{A}_{ki} \mathbf{Q}_h \mathbf{A}_{ki}^H + \mathbf{A}_{ki} \mathbf{Q}_k \mathbf{A}_{ki}^H + \mathbf{A}_{ki} \mathbf{Q}_{\mathcal{Q}_{ki}} \right) &= \left( \mathbf{I} - \sum_{\substack{t \in \mathcal{A}_i \\ t \neq k}} \mathbf{W}_{ti} \mathbf{A}_{ti} - \mathbf{W}_{ii} \right) \mathbf{Q}_h \mathbf{A}_{ki}^H. \end{aligned}$$

Solve the above equation system, the optimal weight combining matrices  $\{\mathbf{W}_{ki}^{opt}, \mathbf{W}_{ii}^{opt}, k \in \mathcal{A}_i\}$  can be derived as:

$$\begin{bmatrix} \mathbf{W}_{ii}^{opt} & \mathbf{W}_{l_1 i}^{opt} & \dots & \mathbf{W}_{l_{C_i} i}^{opt} \end{bmatrix} = \mathbf{Q}_h \Upsilon_i \Omega_i^{-1}.$$

Let

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{ii}^{opt} & \mathbf{W}_{l_1 i}^{opt} & \dots & \mathbf{W}_{l_{C_i} i}^{opt} \end{bmatrix}$$

$$\Theta_i = \begin{bmatrix} \mathbf{Q}_i & 0 & \dots & 0 \\ 0 & \mathbf{P}_{l_1 i} - \mathbf{A}_{l_1 i} \mathbf{Q}_h \mathbf{A}_{l_1 i}^H & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{P}_{l_{C_i} i} - \mathbf{A}_{l_{C_i} i} \mathbf{Q}_h \mathbf{A}_{l_{C_i} i}^H \end{bmatrix},$$

then

$$\Omega_i = \Theta_i + \Upsilon_i^H \mathbf{Q}_h \Upsilon_i.$$

Since

$$\mathbf{W} = \mathbf{Q}_h \Upsilon_i \Omega_i^{-1},$$

the optimum per dimensional MSE satisfies

$$\begin{aligned} D^{(i)opt} &= \frac{1}{n} \text{Tr} ((\mathbf{W} \Upsilon_i^H - \mathbf{I}) \mathbf{Q}_h (\mathbf{W} \Upsilon_i^H - \mathbf{I})^H + \mathbf{W} \Theta_i \mathbf{W}^H) \\ &= \frac{1}{n} \text{Tr} (\mathbf{Q}_h - \mathbf{Q}_h \Upsilon_i \Omega_i^{-1} \Upsilon_i^H \mathbf{Q}_h) \\ &\stackrel{(a)}{=} \frac{1}{n} \text{Tr} (\mathbf{Q}_h^{-1} + \Upsilon_i \Theta_i^{-1} \Upsilon_i^H)^{-1} \\ &= \frac{1}{n} \text{Tr} \left( \mathbf{Q}_h^{-1} + \sum_{k \in \mathcal{A}_i} \mathbf{A}_{ki}^H (\mathbf{P}_{ki} - \mathbf{A}_{ki} \mathbf{Q}_h \mathbf{A}_{ki}^H)^{-1} \mathbf{A}_{ki} + \mathbf{Q}_i^{-1} \right)^{-1} \\ &= \frac{1}{n} \text{Tr} (\mathbf{Q}_h^{-1} + \mathbf{Q}_i^{-1} + \mathbf{\Lambda}_i)^{-1}, \end{aligned}$$

where (a) follows from the Woodbury identity

$$(\mathbf{A} + \mathbf{C} \mathbf{B} \mathbf{C}^H)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{C} (\mathbf{B}^{-1} + \mathbf{C}^H \mathbf{A}^{-1} \mathbf{C})^{-1} \mathbf{C}^H \mathbf{A}^{-1}.$$

This concludes the proof.

## APPENDIX B

## PROOF OF PROPOSITION 3

In order to prove this proposition, we need the following lemma:

*Lemma 1.* for a random vector  $\mathbf{x}$ , if an optimal  $S$  levels Euclidean distance distortion based quantizer applied on a random vector  $\mathbf{y} = \mathbf{B}^{\frac{1}{2}}\mathbf{x}$  has the codebook  $\{\mathbf{y}_1, \dots, \mathbf{y}_S\}$  and associated partition  $\{\mathcal{P}_1, \dots, \mathcal{P}_S\}$ , then the optimal  $S$  level weighted square error distortion based quantizer applied on the random vector  $\mathbf{x}$  with shaping matrix  $\mathbf{B}$  will have the codebook  $\{\mathbf{B}^{-\frac{1}{2}}\mathbf{y}_1, \dots, \mathbf{B}^{-\frac{1}{2}}\mathbf{y}_S\}$  and associated partition  $\{\mathbf{B}^{-\frac{1}{2}}[\mathcal{P}_1], \dots, \mathbf{B}^{-\frac{1}{2}}[\mathcal{P}_S]\}$ , where the  $\mathbf{B}^{-\frac{1}{2}}[\mathcal{P}_i]$  is defined as  $\mathbf{B}^{-\frac{1}{2}}[\mathcal{P}_i] = \{\mathbf{x} : \exists \mathbf{y} \in \mathcal{P}_i \text{ s.t. } \mathbf{x} = \mathbf{B}^{-\frac{1}{2}}\mathbf{y}\}$ .

*Proof:* Consider the distortion associated with the optimal codebook and partition:

$$\begin{aligned} D_{\mathbf{y}} &= \sum_{j=1}^S f_{\mathbf{y}}(\mathbf{y}_j) \int_{\mathbf{y} \in \mathcal{P}_j} (\mathbf{y} - \mathbf{y}_j)^H (\mathbf{y} - \mathbf{y}_j) d\mathbf{y}. \\ D_{\mathbf{x}} &= \sum_{j=1}^S f_{\mathbf{x}}(\mathbf{x}_j) \int_{\mathbf{x} \in \mathcal{M}_j} (\mathbf{x} - \mathbf{x}_j)^H \mathbf{B} (\mathbf{x} - \mathbf{x}_j) d\mathbf{x} \\ &= \det(\mathbf{B}^{-1}) \sum_{j=1}^S f_{\mathbf{x}}(\mathbf{x}_j) \int_{\mathbf{u} \in \mathbf{B}^{\frac{1}{2}}[\mathcal{M}_j]} (\mathbf{u} - \mathbf{B}^{\frac{1}{2}}\mathbf{x}_j)^H (\mathbf{u} - \mathbf{B}^{\frac{1}{2}}\mathbf{x}_j) d\mathbf{u}, \end{aligned}$$

where  $\{\mathbf{x}_1, \dots, \mathbf{x}_S\}$  and  $\{\mathcal{M}_1, \dots, \mathcal{M}_S\}$  denote the optimal codebook and partition for the quantizing  $\mathbf{x}$ . Let  $\mathbf{B}^{\frac{1}{2}}\mathbf{x}_j = \mathbf{y}_j$  and  $\mathbf{B}^{\frac{1}{2}}[\mathcal{M}_j] = \mathcal{P}_j$ , since

$$f_{\mathbf{y}}(\mathbf{y}_j) = f_{\mathbf{x}}(\mathbf{x}_j) \det(\mathbf{B}^{-1}),$$

by change of variable in the integral, we can get  $D_{\mathbf{y}} = D_{\mathbf{x}}$ . ■

Thus, according to Lemma 1,

$$\begin{aligned} \mathbf{Q}_{\mathbf{x}} &= \mathbb{E}\{(\mathbf{x} - \mathcal{Q}(\mathbf{x}))(\mathbf{x} - \mathcal{Q}(\mathbf{x}))^H\} \\ &= \mathbf{B}^{-\frac{1}{2}} \mathbb{E}\{(\mathbf{y} - \mathcal{Q}(\mathbf{y}))(\mathbf{y} - \mathcal{Q}(\mathbf{y}))^H\} \mathbf{B}^{-\frac{H}{2}} \\ &= \mathbf{B}^{-\frac{1}{2}} \mathbf{Q}_{\mathbf{y}} \mathbf{B}^{-\frac{H}{2}}, \end{aligned}$$

where  $\mathbf{y} = \mathbf{B}^{\frac{1}{2}}\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{B}^{\frac{1}{2}}\mathbf{\Gamma}\mathbf{B}^{\frac{H}{2}})$ . Let  $\mathbf{t} = [\Re(\mathbf{y})^T \Im(\mathbf{y})^T]^T$ , then  $\mathbf{t} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Phi})$  and

$$\mathbf{\Phi} = \frac{1}{2} \begin{bmatrix} \Re(\mathbf{B}^{\frac{1}{2}}\mathbf{\Gamma}\mathbf{B}^{\frac{H}{2}}) & \Im(\mathbf{B}^{\frac{1}{2}}\mathbf{\Gamma}\mathbf{B}^{\frac{H}{2}}) \\ \Im(\mathbf{B}^{\frac{1}{2}}\mathbf{\Gamma}\mathbf{B}^{\frac{H}{2}}) & \Re(\mathbf{B}^{\frac{1}{2}}\mathbf{\Gamma}\mathbf{B}^{\frac{H}{2}}) \end{bmatrix}.$$

Furthermore, it is proven in [34] that

$$\mathbf{Q}_{\mathbf{t}} = \mathbb{E}\{(\mathbf{t} - \mathcal{Q}(\mathbf{t}))(\mathbf{t} - \mathcal{Q}(\mathbf{t}))^T\} = D_{\mathbf{t}}\mathbf{I}_n,$$

where the average distortion  $D_{\mathbf{t}} = \frac{1}{n} \text{Tr}(\mathbf{Q}_{\mathbf{t}})$  is obtained from [32] for large  $S$  as

$$\begin{aligned} D_{\mathbf{t}} &= S^{-\frac{2}{n}} M_n \left( \int f_{\mathbf{t}}(\mathbf{t})^{\frac{n}{n+2}} d\mathbf{t} \right)^{\frac{n+2}{n}} \\ &= S^{-\frac{2}{n}} M_n 2\pi \left( \frac{n+2}{n} \right)^{\frac{n}{2}+1} \det(\Phi)^{\frac{1}{n}}, \end{aligned}$$

where  $f_{\mathbf{t}}(\cdot)$  is the probability density function (p.d.f) of  $\mathbf{t}$ . Finally, from the expression of  $\mathbf{Q}_{\mathbf{t}}$  and by a real-to-complex conversion, we get

$$\mathbf{Q}_{\mathbf{y}} = 2S^{-\frac{1}{n}} M_{2n} 2\pi \left( \frac{n+1}{n} \right)^{n+1} \det(\Phi)^{\frac{1}{2n}} \mathbf{I}_n,$$

which leads to

$$\begin{aligned} \mathbf{Q}_{\mathbf{x}} &= 2S^{-\frac{1}{n}} M_{2n} 2\pi \left( \frac{n+1}{n} \right)^{n+1} \det(\Phi)^{\frac{1}{2n}} \mathbf{B}^{-1} \\ &= S^{-\frac{1}{n}} M_{2n} 2\pi \left( \frac{n+1}{n} \right)^{n+1} \det(\mathbf{B})^{\frac{1}{n}} \det(\Gamma)^{\frac{1}{n}} \mathbf{B}^{-1} \\ &= \mathbf{Q}_0^{(S)}(\Gamma) \det(\mathbf{B})^{\frac{1}{n}} \mathbf{B}_{ki}^{-1}, \end{aligned}$$

where

$$\mathbf{Q}_0^{(S)}(\Gamma) = S^{-\frac{1}{n}} M_{2n} 2\pi \left( \frac{n+1}{n} \right)^{n+1} \det(\Gamma)^{\frac{1}{n}} \mathbf{I}_n.$$

## APPENDIX C

### PROOF OF PROPOSITION 4

We first give some necessary lemmas:

*Lemma 2.* [35, Lemma 2.5] The matrix function  $h : \mathbf{X} \mapsto -\mathbf{X}^{-1}$  is both a strictly matrix concave function and a matrix monotone increasing function.

*Lemma 3.* If  $\mathbf{X}$  is positive semi-definite hermitian matrix, then matrix function  $h : \mathbf{X} \mapsto \text{Tr}(\mathbf{X}^{-1})$  is convex and matrix monotone decreasing.

*Proof:* Consider  $g(t) = h(\mathbf{Z} + t\mathbf{V})$ , where  $\mathbf{Z}$  is positive semi-definite hermitian matrix,  $\mathbf{V}$  is hermitian matrix, since

$$\frac{d^2}{dt^2} g(t)|_{t=0} = \text{Tr} \left( 2(\mathbf{Z}^{-1}\mathbf{V})\mathbf{Z}^{-1}(\mathbf{Z}^{-1}\mathbf{V})^{-1} \right) \geq 0,$$

therefore  $f(\mathbf{X})$  is convex.

For  $\mathbf{X} \succeq \mathbf{Y} \succeq \mathbf{0}$ , according to Lemma 2,

$$\mathbf{Y}^{-1} \succeq \mathbf{X}^{-1} \succeq \mathbf{0}$$

$$h(\mathbf{X}) - h(\mathbf{Y}) = \text{Tr}(\mathbf{X}^{-1} - \mathbf{Y}^{-1}) \leq 0,$$

therefore, the function is matrix monotone decreasing. ■

Adopt the notations in Proposition 2, we will prove the convexity of problem (12). Since

$$\det(\mathbf{B}_{ki}) = 1,$$

according to (11),

$$\mathbf{Q}_{\mathcal{Q}_{ki}} = \mathbf{Q}_0^{(S)}(\Gamma)\mathbf{B}_{ki}^{-1}.$$

Let

$$f = \mathbf{Q}_{\mathbf{h}}^{-1} + \mathbf{Q}_i^{-1} + \Lambda_i,$$

with  $\Lambda_i$  defined in Corollary 2. Consider the convexity of composite function: if function  $h$  is concave and matrix monotone increasing, function  $g$  is a concave function, then the composite function  $h \circ g$  is concave. According to Lemma 2,

$$h = -\mathbf{X}^{-1},$$

is matrix concave and matrix monotone increasing for  $\mathbf{X}$ . According to Lemma 2,

$$g = \mathbf{A} - \mathbf{B}_{ki}^{-1},$$

with constant matrix  $\mathbf{A}$  is concave for  $\mathbf{B}_{ki}$ . Therefore the functions

$$\Lambda_i = h \circ g$$

$$f = \mathbf{Q}_{\mathbf{h}}^{-1} + \mathbf{Q}_i^{-1} + \Lambda_i$$

are concave for  $\mathbf{B}_{ki}$ .

Consider the convexity of composite function: if function  $h$  is convex and matrix monotone decreasing, function  $g$  is concave function, then the composite function  $h \circ g$  is convex. According to Lemma 3,

$$h = \text{Tr}(\mathbf{X}^{-1})$$

is convex and matrix monotone decreasing for  $\mathbf{X}$ , since

$$g = \mathbf{Q}_h^{-1} + \mathbf{Q}_i^{-1} + \mathbf{\Lambda}_i$$

is concave, therefore we can conclude that

$$D^{(i)opt} = h \circ g$$

is convex for  $\mathbf{B}_{ki}$ .

When the coordination link rate  $R_{ki}$  is sufficiently large, use twice matrix inverse approximation:

$$\begin{aligned} D^{(i)opt} &= \frac{1}{n} \text{Tr} \left[ \mathbf{Q}_h^{-1} + \mathbf{Q}_i^{-1} + \sum_{k \in \mathcal{A}_i} ((\mathbf{Q}_h + \mathbf{Q}_k)(\mathbf{Q}_h + \mathbf{Q}_k - \mathbf{Q}_{\mathcal{Q}_{ki}})^{-1}(\mathbf{Q}_h + \mathbf{Q}_k) - \mathbf{Q}_h)^{-1} \right]^{-1} \\ &\stackrel{(a)}{\simeq} \frac{1}{n} \text{Tr} \left( \sum_{k \in \mathcal{A}_i} (\mathbf{Q}_k + \mathbf{Q}_{\mathcal{Q}_{ki}})^{-1} + \mathbf{Q}_i^{-1} + \mathbf{Q}_h^{-1} \right)^{-1} \\ &\stackrel{(b)}{\simeq} \frac{1}{n} \text{Tr} \left( \sum_{k \in \mathcal{A}_i} (\mathbf{Q}_k^{-1} - \mathbf{Q}_k^{-1} \mathbf{Q}_{\mathcal{Q}_{ki}} \mathbf{Q}_k^{-1}) + \mathbf{Q}_i^{-1} + \mathbf{Q}_h^{-1} \right)^{-1}, \end{aligned}$$

where (a) and (b) follows from the matrix inverse first order approximation: if  $\mathbf{X}^n \rightarrow 0$  when  $n \rightarrow \infty$ , then

$$(\mathbf{A} + \mathbf{X})^{-1} \simeq \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{X} \mathbf{A}^{-1}.$$

Similar to the previous analysis, according to convexity for composite matrix function, apply Lemma 3 and Lemma 2, we can conclude the above approximation for  $D^{(i)opt}$  is convex.

If we relax the constraint  $\det(\mathbf{B}_{ki}) = 1$  to  $\det(\mathbf{B}_{ki}) \geq 1$ , the feasible set for  $\mathbf{B}_{ki}$  is therefore a convex set and the optimization for this relaxed problem is minimizing a convex function over a convex set, which is a convex optimization problem and the optimal  $\mathbf{B}_{ki}$  should attain at the boundary  $\det(\mathbf{B}_{ki}) = 1$ . This concludes the proof.

#### ACKNOWLEDGMENT

The authors would like to thank Dr. Paul de Kerret for the fruitful discussions and valuable suggestions to improve the work.



## REFERENCES

- [1] P. Marsch and G. P. Fettweis, *Coordinated Multi-Point in Mobile Communications: from theory to practice*. Cambridge University Press, 2011.
- [2] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, “Multi-cell MIMO cooperative networks: A new look at interference,” *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, 2010.
- [3] S. Venkatesan, A. Lozano, and R. Valenzuela, “Network MIMO: Overcoming intercell interference in indoor wireless systems,” in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC)*, 2007.
- [4] M. Karakayali, G. Foschini, and R. Valenzuela, “Network coordination for spectrally efficient communications in cellular systems,” *IEEE trans. Wireless Commun.*, vol. 13, no. 4, pp. 56–61, 2006.
- [5] H. Dahrouj and W. Yu, “Coordinated beamforming for the multicell multi-antenna wireless system,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, 2010.
- [6] R. Zakhour and D. Gesbert, “Distributed multicell-MISO precoding using the layered virtual SINR framework,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2444–2448, 2010.
- [7] E. Bjornson, G. Zheng, M. Bengtsson, and B. Ottersten, “Robust monotonic optimization framework for multicell MISO systems,” *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2508–2523, 2012.
- [8] V. R. Cadambe and S. A. Jafar, “Interference alignment and degrees of freedom of the K-user interference channel,” *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, 2008.
- [9] A. Sanderovich, O. Somekh, and S. Shamai, “Uplink macro diversity with limited backhaul capacity,” in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2007.
- [10] S. Shamai, O. Simeone, O. Somekh, and H. Poor, “Joint multi-cell processing for downlink channels with limited-capacity backhaul,” in *Proc. Information Theory and Applications Workshop*, 2008.
- [11] P. Marsch and G. Fettweis, “On base station cooperation schemes for downlink network MIMO under a constrained backhaul,” in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2008.
- [12] R. Bhagavatula and R. W. Heath, “Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems,” *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 800 – 811, 2011.
- [13] M. Rezaee, M. Guillaud, and F. Lindqvist, “CSIT sharing over finite capacity backhaul for spatial interference alignment,” in *Proc. IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2013.
- [14] Q. Li, D. Gesbert, and N. Gresset, “A cooperative channel estimation approach for coordinated multipoint transmission networks,” in *Proc. IEEE International Conference on Communication Workshop (ICCW)*, 2015.
- [15] —, “Joint precoding over a master-slave coordination link,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.
- [16] M. Fleming, Q. Zhao, and M. Effros, “Network vector quantization,” *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1584–1604, 2004.
- [17] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.
- [18] A. Del Coso and S. Simoens, “Distributed compression for the uplink of a backhaul-constrained coordinated cellular network,” *arXiv preprint arXiv:0802.0776*, 2008.

- [19] H. Yamamoto, "Source coding theory for multiterminal communication systems with a remote source," *IEICE TRANS-ACTIONS (1976-1990)*, vol. 63, no. 10, pp. 700–706, 1980.
- [20] S. C. Draper, "Successive structuring of source coding algorithms for data fusion, buffering, and distribution in networks," Ph.D. dissertation, Massachusetts Institute of Technology, 2002.
- [21] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.
- [22] Z. Liu, S. Cheng, A. D. Liveris, and Z. Xiong, "Slepian-wolf coded nested lattice quantization for wyner-ziv coding: high-rate performance analysis and code design," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4358–4379, 2006.
- [23] D. Rebollo-Monedero, S. Rane, and B. Girod, "Wyner-ziv quantization and transform coding of noisy sources at high rates," in *Proc. Asilomar Conference on Signals, Systems and Computers (ACSSC)*, 2004.
- [24] R. Fritzsche and G. Fettweis, "Distributed robust sum rate maximization in cooperative cellular networks," in *Proc. IEEE International Conference on Communications Workshops (ICCW)*, 2013.
- [25] S. Wagner, R. Couillet, M. Debbah, and D. T. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.
- [26] P. de Kerret and D. Gesbert, "Degrees of freedom of the network MIMO channel with distributed CSI," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6806–6824, 2012.
- [27] I. Jee and R. Haddaad, "Optimum design of vector-quantized subband codecs," *IEEE Trans. Signal Process.*, vol. 46, no. 8, pp. 2239–2243, Aug 1998.
- [28] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.
- [29] P. Marsch and G. Fettweis, "On downlink network MIMO under a constrained backhaul and imperfect channel knowledge," in *Proc. Global Telecommunications Conference ( GLOBECOM)*, 2009.
- [30] P. H. Westerink, J. Biemond, and D. E. Boeke, "Scalar quantization error analysis for image subband coding using QMFs," *IEEE Trans. Signal Process.*, vol. 40, no. 2, pp. 421–428, 1992.
- [31] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.
- [32] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 139–149, Mar 1982.
- [33] J. Conway and N. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 211–226, Mar 1982.
- [34] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, Jul 1996.
- [35] J. Brinkhuis, Z.-Q. Luo, and S. Zhang, "Matrix convex functions with applications to weighted centers for semidefinite programming," *Econometric Institute Research Papers*, Tech. Rep., 2005.